

The Application of Neural Systems in Vibrodiagnosis

TOMASZ ROMANIUK

Institute of Computer Science, Jagiellonian University
Prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland
e-mail: *tomasz.romaniuk@gmail.com*

Abstract. Vibrodiagnosis helps in detecting incipient faults in rotating machines like pumps and generators. Early detection prevents from undesired breakdown of the machine and allows to schedule maintenance times. The application of neural networks in classification of the rotating machine condition has been described in this work. Different types of networks and methods on feature extraction was described and compared. Additionally it was proposed a novelty feature set consisted of harmonics from vibration spectrum. The set were combined with using of probabilistic neural networks which has been modified that it could recognize defects that did not occur in the training set. Such architecture was tested in detection of two defects, shaft misalignment and mass unbalance. It was found that such network works better than a multi layered perceptron with statistical features.

Keywords: neural networks, probabilistic neural networks, vibration monitoring, diagnostic systems.

1. Introduction

Artificial neural networks have been widely developed in the past twenty years. They were found to be a good tool for solving many problems related to function approximation, prediction, classification, pattern recognition and

Tab. 1. An example of vibration severity chart ISO 10816 [1]

Vibration Velocity IPS (RMS)	Group 4 Integrated Driver		Group 3 External Driver		Group 2 Motors $160 \leq H < 315$ mm		Group 1 Motors $H \geq 315$ mm	
	Rigid	Flexible	Rigid	Flexible	Rigid	Flexible	Rigid	Flexible
0.71								
0.43								
0.28								
0.18								
0.14								
0.11								
0.09								
0.06								
0.03								

many others. But still the biggest challenge is adaptation of ANNs to real life applications and such branches as engineering. One of such domain is vibration monitoring which helps in detecting incipient faults in rotating machines like pumps and generators. Early detection prevents undesired breakdown of the machine and allows to schedule maintenance times. Thus, a good detection system improves reliability of the equipment and the whole plant.

2. Vibration monitoring

2.1. Overview

As it was mentioned before, vibration monitoring helps to determine the condition of the machine. It can help in detecting wear of machine elements, bad configuration which can cause such a wear, and also defects. Thanks to this, the operator of the machine can schedule maintenance and replace worn out elements before breakdown. It is noteworthy that modern methods of vibration analysis enable not only anomaly detection but they can also tell why this anomaly occurs. They can find the reason behind the improper working of the machine which is very helpful in the maintenance process. There is a wide range of rotating machines defects. The most common are bearing defects, unbalance and misalignment. In this work we investigated detection of two conditions: shaft misalignment and mass unbalance.

One of the first methods used in machine condition monitoring were vibration severity charts – see Fig. 1 as an example. This chart contains levels for four types of machines with rigid and flexible clutches. The colors on the chart correspond to conditions of the machine. ISO 10816-3 separates the working conditions into four zones:

1. Zone A Green: Vibration values from machines just put into operation.
2. Zone B Yellow: Continuous operation without any restrictions.
3. Zone C Orange: Condition is acceptable only for a limited period of time.
4. Zone D Red: Dangerous vibration values – damage could occur at any time.

We can see that this method is based on measuring general vibration level and comparing it to the severity chart which contains admissible vibration levels for different types of machines. Thus, before diagnosis we must know our machine type as well as operating speed during measurement. Using such a chart one may assess how bad the vibration might be on his machine. Unfortunately we do not know the reasons for increased vibrations, we can only tell if the condition is bad or very bad. When we take a time series representing a vibration measurement then in fact it is a sum of vibrating machine parts. Additionally, such a measurement may often contain different noises, which come from nearby equipment. However, there are some methods for characterization of such measurement. The most common is defining the significance of signal amplitude.

- Peak – it is the distance from the equilibrium point to the biggest peak.
- Peak-to-Peak – it is the distance from the lower peak to the upper peak, when a signal is symmetrical then it is equal to twice value of the Peak.
- RMS – Root Mean Square, it is expressed by Eq. (24), where x_i are samples of discrete signal.
- AVG – arithmetic average, it is rarely used in vibrodiagnosis.

Other useful information that can be extracted from time series is a crest factor which is calculated from the peak amplitude of the signal divided by the RMS value. Common in vibrodiagnosis is also calculating such statistical parameters like skewness (27) and kurtosis (26).

All this parameters are very useful in detecting machine faults, however in many cases this is not sufficient to full detection. The method which changed

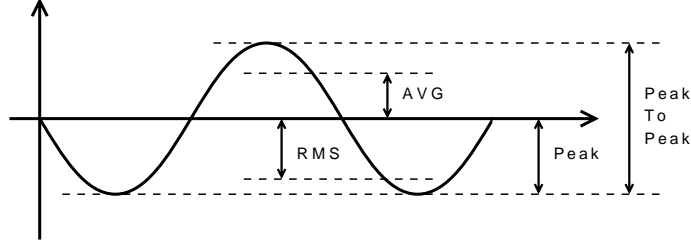


Fig. 1. Different methods of measuring signal amplitude

this and became a breakthrough in vibration monitoring was spectral analysis. A spectrum is a graphical display of a signal in frequency domain, on such plot we can obtain amplitudes for particular frequencies. It is known that signal in frequency domain contains the same information as in time domain, but vibration analysis which uses first form is much easier. A large number of rotation machinery defects manifest itself in a spectrum in form of single peaks. The frequencies and amplitudes of such peaks depend on the type and severity of diagnosed defect and constitute an expert knowledge. A common practice is doing measurements on various locations or directions and comparing them. Horizontal vibrations give different information about state of the machine than vertical or axial. Additionally when we have two measurements from distinct locations but taken at the same time we can analyze a phase shift between them. For example, a misalignment is characterized by high axial vibration and 180° phase value across the coupling. Also magnitudes of the second and third harmonics should be increased. On the Fig. 2 we have an example of horizontal (2a) and axial (2b) vibration of the machine with misalignment, we can clearly recognize the increased vibrations in 2x, 3x, 4x and next harmonics especially at axial measurement. The green circles represent this harmonics. For better analysis, vibration signal can be filtered before computing the spectrum. Low-pass and high pass filters are used for removing high and low-frequency components respectively. Another proven method is using signal demodulation for detecting bearing's defects. Demodulation makes the diagnostic process a little more independent of a particular machine since it focuses on the low-amplitude high-frequency broadband signals characterizing the bearing condition [5].

2.2. AI systems

Artificial intelligence systems have been applied to many problems in industry. They can speed up a complicated process of diagnosis and often

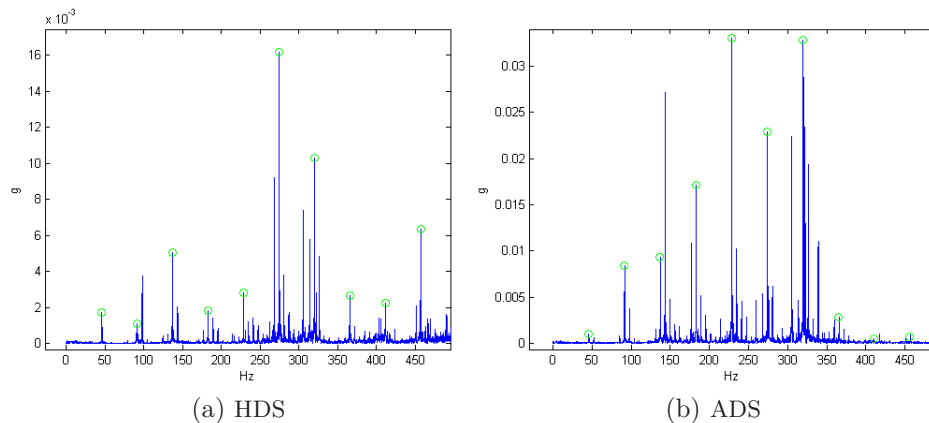


Fig. 2. An example of misalignment

they can work without any expert personnel. As we could see, vibration analysis has many rules and methods which could be formalized and used for building an expert system. Especially the spectral information seemed to be the most useful in such systems. As an example, we briefly present an expert system which was developed at James Cook University in Australia [3]. The database of this system is represented by 75 rules in form of: **If** condition **Then** rule1 or diagnosis1 **Else** rule2 or diagnosis2. Starting from the base rules they create a decision tree. Nodes of the tree contain conditions and leafs represent a particular diagnosis. As an input data were used triaxial spectra and time series, also demodulated vibration measurements were taken into consideration while designing diagnostic rules. Most of them check a presence of characteristic peaks in the spectrum and also amplitude of these peaks. Earmarked for this is a special algorithm, which first normalizes spectrum in relation to the biggest peak, and then looks for peaks using thresholds. These thresholds depend on specific machine and should be set up by maintenance engineers. They have percentage values which enable correct operation even with lack of historical data. Additionally, except from the diagnosis the system is computing a confidence factor that represents resemblance of detected peak to the theoretical peak. Value of it is calculated by using linear fuzzy operators, and gives to the operator an information about accuracy of the diagnosis.

There are also other expert systems dedicated to vibration analysis. In 2005 Bo-Suk Yanga, Dong-Soo Limb, Andy Chit Chio Tan created VIBEX [10], which is also based on spectral analysis. In contrast to the system developed at the University of Townsville, its diagnostic engine uses decision trees and decision tables based on the cause-symptom matrix and

Bayesian algorithm. It is also worth to mention that there exist commercial systems as ExpertAlert created by American company DLI Engineering dealing with vibrodiagnosis since 1988. Its database contains over 4,500 diagnostic rules and 650 machine fault messages. It shows how much knowledge and time is needed to create a real, good-working expert system.

The use of Artificial Neural Networks in detecting machine bad condition is quite common. One of the first publication on this subject was the work [4] where a three-layered perceptron was used for classification of four machine conditions: no fault, unbalance, increased rub, unbalance and rub at the same time. The crucial aspect of the work was comparing different methods of feature extraction from time series. There were computed central moments (1) that characterize the probability density function of time series, because they are time-invariant and give an indication of the machine state [4].

$$\mu_i = \frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{X})^i, \quad (1)$$

where \bar{X} is a mean of the vector X .

It was proved that helpful in the signal characterization is the use of differentiation and integration of the time series. They can be approximated by the equations (2) and (3). The differentiation amplifies the high frequencies and attenuates the low frequencies, therefore it is easier to recognize faults that manifest in high-frequency range. The integration works inversely. Moments of differential and integral were attached to the network input vector.

$$\frac{dx}{dt} \approx d(n) = x(n) - x(n-1), \quad (2)$$

$$\int x(t)dt \approx i(n) = x(n) + i(n-1). \quad (3)$$

It was noticed that vertical and horizontal measurements can be used to create complex time series:

$$z(t) = x(t) + jy(t), \quad (4)$$

where $x(t)$ and $y(t)$ are respectively horizontal and vertical time series. Such complex variable represents the movement of shaft on the plane and it was found useful in better classification. The best features set achieved over 99.43% correct classifications.

An extension of the previous method was publications [6, 11], where a selection of features based on genetic algorithm were investigated. Except from moments, cumulants were computed, which can be expressed by using

central moments:

$$k_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} k_k \mu'_{n-k}. \quad (5)$$

Additionally the vibration signals were passed through a low-pass filter and a high-pass filter which give a similar effect to differentiation and integration. Also a 32-point FFT of time series were attached to the network input vector. It was found that the best configuration varied from 6 to 12 features and gave 100% correct classifications.

The number of various machine faults is very large and additionally they can occur simultaneously which makes the diagnostic process more difficult. A neural network which have been learnt to detect misalignment and unbalance would have problems with proper classification if they would occur at the same time. We can obviously teach our network to recognize double and triple faults but this will increase the number of classes to learn. And it is known the more classes the network learns, the weaker classification ability of the network [12]. Therefore, to overcome this problem a decomposition of the problem into various networks organized in a hierarchical order [12] was proposed. Networks from first layer were responsible for detecting single faults, second layer for double and third for triple. Additionally, double and triple faults were divided into subsets to minimize the number of classes to learn.

An interesting aspect was the usage of ellipsoidal neurons. The input of such a neuron is a value of $E(X)$ function, expressed by:

$$E(X) = -(X - M)^T (D^T D)^{-1} (X - M) + 1, \quad (6)$$

where D is a diagonal matrix, equation $E(X) = 0$ defines the n -dimensional ellipsoid in the center M . When a value of the activation function is in the range $(0.5, 1)$ it means that the input vector is inside the hyperellipsoid defined by the neuron, otherwise it is outside of it. It is an important feature which enables to detect faults that did not occur in the training set.

Another type of artificial networks applied to vibrodiagnosis was Self Organizing Map (SOM). SOM maps multi-dimensional data into one or two dimensional grid that represents topological properties of the input pattern. Thus it is used rather in data analysis than classification problems. The work [2] proved that analysis of such maps can be useful in vibration analysis. It is possible to track dependencies between location of the various parameters used in training and fault occurrences. For example from the map we can clearly see that unbalance manifests in higher amplitude of the first harmonic. There is also possibility of connecting the self-organizing layer with a linear layer. A supervised learning has to be applied to the linear layer in order

to assign the neurons to corresponding classes. The architecture of such a network is called the LVQ network (Learning Vector Quantification) and has been used in machine fault detection with success [9].

3. PPNs applications to vibration monitoring

3.1. Theoretical model

Models of Probabilistic Neural Networks (PPNs) were proposed by D.F. Specht [7, 8]. They had been found very useful in classification problems and were adopted in engineering applications. From the mathematical point of view they are an implementation of Bayesian classifier with a non-parametric kernel estimation.

THEOREM 1 (Bayes law). *Given a probability space (Ω, Σ, P) and events $A_1, \dots, A_n \in \Sigma$ that satisfied the conditions:*

1. $P(A_i) > 0$ for any $i = 1, \dots, n$.
2. $A_i \cap A_j = \emptyset$, for every $i \neq j$.
3. $A_1 \cup \dots \cup A_n = \Omega$.

Then for every event $B \in \Sigma$ there is the following equality

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}, \quad (7)$$

for every $k = 1, \dots, n$.

THEOREM 2 (Law of total probability). *Let the conditions from Theorem 1 be satisfied. Then for every event $B \in \Sigma$ there is the following equality*

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i). \quad (8)$$

These two theorems are a basis for derivation of the naive Bayes classifier. Let $x \in \mathbb{R}^d$ be d-dimensional pattern, which is associated with a class $r \in \{C_1, C_2, \dots, C_k\}$, where k is the number of possible classes. The probability

that x belongs to the class C_r can be written with using the Bayes law (7) in the following way

$$p(C_r|x) = \frac{p(x|C_r)p(C_r)}{\sum_{i=1}^k p(x|C_i)}. \quad (9)$$

Then taking equation (8) we simplify denominator to

$$p(C_r|x) = \frac{p(x|C_r)p(C_r)}{p(x)}. \quad (10)$$

We are interested in finding the class r with the biggest probability $p(C_r|x)$. But in fact we are finding the biggest denominator because $p(x|C_i)$, $i = 1, 2, \dots, k$ are probabilities *a priori* of class occurrences and $p(x)$ is a constant.

$$r = \arg \max_i p(x|C_i)p(C_i), \quad \text{for } i = 1, 2, \dots, k. \quad (11)$$

This is the formula for the Bayes classifier and it is known as the maximum a posteriori or MAP decision rule. But still we do not know $p(x|C_i)$ probabilities, which are necessary for finding the maximum. Thus we will use a kernel estimator of a probability density function.

DEFINITION 3. *The function $K: \mathbb{R}^d \rightarrow \mathbb{R}$ is called a kernel, when:*

1. $\int_{\mathbb{R}^d} |K(x)|^2 dx < \infty$
2. $\int_{\mathbb{R}^d} K(x) dx = 1$
3. $K(-x) = K(x)$, *dla* $x \in \mathbb{R}^d$.

An example of kernels are the following functions:

- **Gaussian:**

$$K(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|^2}{2}\right), \quad (12)$$

- **Epanechnikov:**

$$K(x) = \begin{cases} \frac{3}{4}(1 - \|x\|^2), & \|x\| < 1 \\ 0, & \|x\| \geq 1 \end{cases}, \quad (13)$$

- **Triangle:**

$$K(x) = \begin{cases} 1 - \|x\|, & \|x\| < 1 \\ 0, & \|x\| \geq 1 \end{cases}. \quad (14)$$

Now we can formulate the definition of the kernel estimator.

DEFINITION 4. *Let x_1, \dots, x_n be a sample from continuous distribution with density function f . For a given kernel K and the number $h > 0$ which is called a smoothing parameter (it controls the size of the kernel) we define the kernel estimator of density f as:*

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad \text{for } x \in \mathbb{R}^d. \quad (15)$$

In another way a kernel estimator approximates (estimates) a probability density function by using the statistical sample. In probabilistic networks the most common is using Gaussian kernels. Then as a smoothing parameter h it is used the width of the Gaussian function σ . Taking the formula (12) and inserting into (15) we have

$$\hat{f}(x) = \frac{1}{(2\pi\sigma^2)^{d/2}N} \sum_{i=1}^N \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right). \quad (16)$$

Now getting back to the formula (11) we derive the probability $p(x|C_i)$ by using such estimator. Our statistical sample will be all patterns from the class C_i

$$\hat{p}(x|C_i) = \frac{1}{(2\pi\sigma^2)^{d/2} |C_i|} \sum_{j=1}^{|C_i|} \exp\left(-\frac{\|x - x_j\|^2}{\sigma^2}\right), \quad (17)$$

where $|C_i|$ is a quantity of the set C_i . The final form of the Bayes classifier is

$$r = \arg \max_i \frac{1}{(2\pi\sigma^2)^{d/2} |C_i|} \sum_{j=1}^{|C_i|} \exp\left(-\frac{\|x - x_j\|^2}{\sigma^2}\right) p(C_i), \quad \text{for } i = 1, 2, \dots, k. \quad (18)$$

3.2. Network structure

The probabilistic networks used for classification consist of four layers – see Fig. 3. The first one is a classical input layer. The second one is a radial layer with Gaussian kernel functions. The number of neurons in this layer is equal to the number of patterns in the training set. These patterns are centers of Gaussian functions in individual neurons. Thanks to that, the radial layer returns a vector with levels of similarity of the input vector to remembered patterns. The next layer is linear and it realizes the summation of probabilities

for each class, thus the number of neurons in it is always the same as the number of possible classes. An individual neuron in this layer performs such summation for the class associated with it. The values of neurons are set according to the formula (21). The last layer is competitive layer and is responsible for choosing a class with the biggest sum of probabilities. Given an input vector $x \in \mathbb{R}^d$ and training patterns x_1, x_2, \dots, x_n that each of them is associated with one of the class (C_1, C_2, \dots, C_k) , we can write computations of each layer by

$$net_{1,i} = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right), \quad (19)$$

$$net_{2,i} = \sum_j w_{ij} y_j, \quad \text{where } y_j = net_{1,j}, \quad (20)$$

$$w_{i,j} = \begin{cases} 1, & \text{for } x_j \in C_j \\ 0, & \text{for } x_j \notin C_j \end{cases}, \quad (21)$$

$$net_3 = \arg \max_r net_r. \quad (22)$$

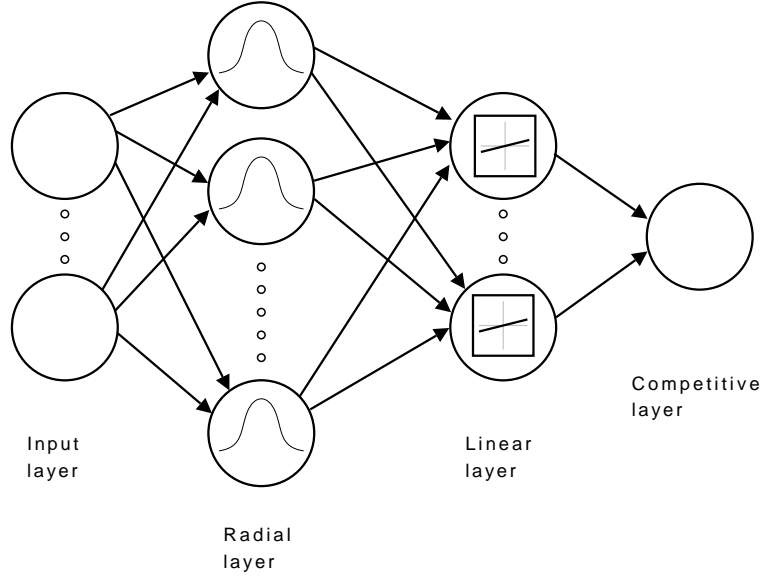


Fig. 3. The structure of probabilistic neural network

However, while constructing a network for machine fault detection, it is impossible to have representative training sets for every defect because the number of them can be very large and simulation of them involves using

a specialized machine. Thus it is possible to change the way the network works. In a situation when a sum of probabilities is very low for all classes we can inform of detection of a new defect. Then such network will have an ability to recognize defects that did not occur in the training set, which is a big advantage over a multi-layered perceptron. The sensitivity of detecting new faults, that is, the limit of probability, could be assigned and changed by maintenance engineers.

3.3. Feature extraction

In the works on the application of neural networks in classification of the machine condition the most common approach to feature extraction was taking statistical features [4, 6, 5]. But it was widely recognized that detection of such faults like misalignment and unbalance is based on analysis of low frequencies, especially on the firsts harmonics. Thus we decided of using it as a feature set. The formal definition of the n -th harmonic can have the following form

DEFINITION 5. *Let $H(\omega)$ be a magnitude of the vibration spectrum at ω frequency and f_r be an operating speed of the machine. The value $H(kf_r)$, for $k = 1, 2, \dots$ will be called a k -th harmonic and denoted by s_k or kx and the value $H(f_r)$ will be called a main harmonic.*

ABBREVIATION 6. *The vector in the form s_1, \dots, s_k will be denoted by S_k .*

The test rig which is described in detail in the next chapter () has the ability of measuring the speed of the motor. However, it has been noticed that this measurement differed from the real speed. This is clearly seen on Fig. (4a) where first ten harmonics computed from measured speed were marked by red circles. We can see how disproportion between circles and real harmonics represented by significant peaks is increasing with frequency. To eliminate this a peak correction algorithm consisting of two steps was used. The first is finding the biggest peak in the neighborhood of measured speed. If the neighborhood will be too small we don't achieve the fundamental harmonic if it will be too big we can recognize a harmonic sideband. However, resolution of the spectrum is limited by the FFT algorithm and such peak also often can slightly differ from the real one. Thus the second step is checking if next harmonics computed from our founded peak overlay with peaks in the given spectrum. If not the fundamental speed must be corrected. The effect

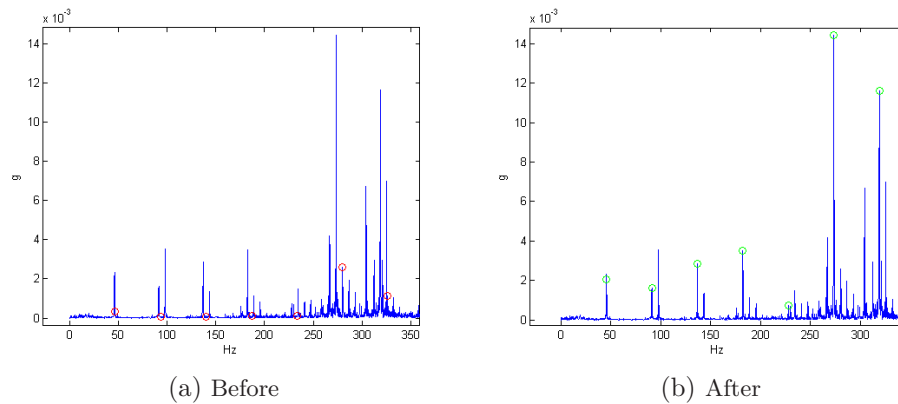


Fig. 4. Peaks correction

of a such correction is depicted in Fig. (4b) where harmonics were marked by green circles.

4. Results

4.1. Training set

The data used for that work was obtained from simulation station which consisted of electric motor (AC, 60W) connected with a shaft through a clutch – see Fig. 5. The station enables a simulation of unbalance and misalignment at different levels. The measurements were taken by using DLI Watchman SpriteMAX, which enables measurements at four locations at the same time. Measurements were taken in three different locations:

1. HDS – horizontal drive side.
2. HNDS – horizontal non-drive side.
3. ADS – axial drive side.

Unbalance was simulated by attaching additional screws to the shaft. All types of unbalance depicts Tab. 2. It should be noticed that situations E and F are in fact good conditions because screws were attached symmetrically and the shaft mass were well-balanced.

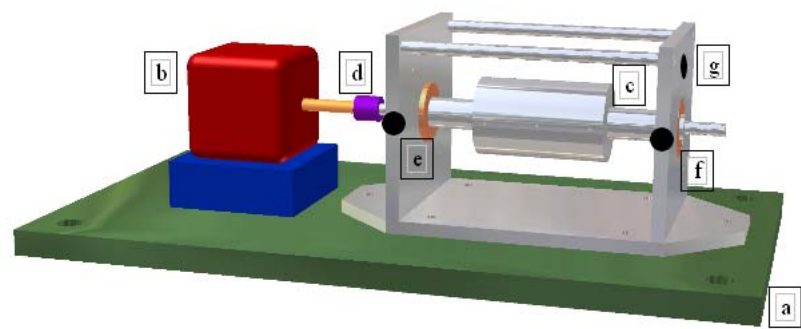
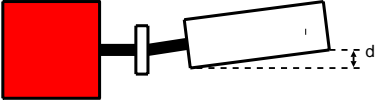
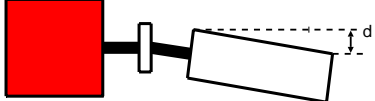
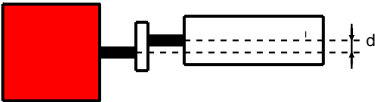


Fig. 5. The simulation station: a) solid basis, b) engine, c) shaft, d) clutch, e) location 1, f) location 2, g) location 3

Tab. 2. Misalignment types

Signature		Signature	
A		F	

Tab. 3. Unbalance types

Type	Signature	
Angular foreword	AFd	
Angular backward	ABd	
Parallel	Pd	

Tab. 4. All simulated conditions

Defect	Signature	Count
normal	ok	27
	ok (E)	9
	ok (F)	9
unbalance	A	6
	B	6
	C	6
	D	6
	G	6
	H	6
	J	6
misalignment	AF1	6
	AF1	6
	AF2	6
	AF3	6
	AF4	6
	AB1	6
	AB2	6
	AB3	10
	P1	6
	P2	6
	P3	6

The cage had ability of moving backward and foreword which enabled of testing different types of misalignment. There were simulated angular and parallel misalignment with different levels of intensity. The Tab. 3 shows how such misalignment looks from the top view of the test rig. All simulated situations were listed in Tab. 4.

Every measurement was taken at 2.5 kHz frequency and was 2 seconds long. In practice we have given time series of 12,500 samples each. In order to increase a number of training samples, each pattern was divided into a three non-overlapping parts. This is a common practice and it has been used in works [4, 5]. One of these parts was included in the testing set and others in the training set. Additionally over 30% of the patterns were generated from existing time series by adding random noise with normal distribution and a standard deviation $\sigma = 1$. The value of noise was about 1–5% of the signal value depending on operating speed and the simulated defect. The measurements were taken at three different speeds: 37, 42, 47 Hz.

Tab. 5. The division of training and testing sets

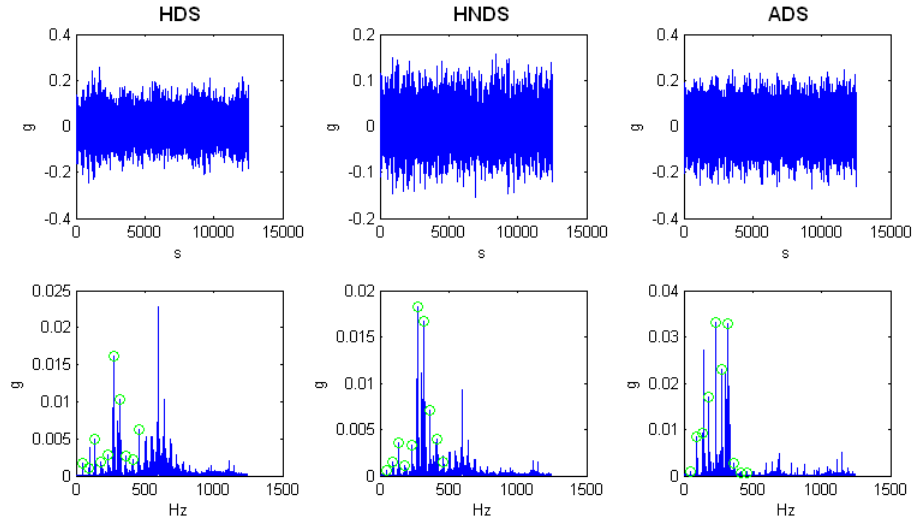
	Training set	Testing set
OK	90	45
MIS	128	64
UNB	84	42

Additionally misalignment and unbalance were simulated in different degree of intensity. The training set obtained in such a way had 302 examples and the testing set had 151. The precise division is contained in the Tab. 5. For better insight into the problem Fig. 6 contains representative time waveform and FFT of three tested conditions.

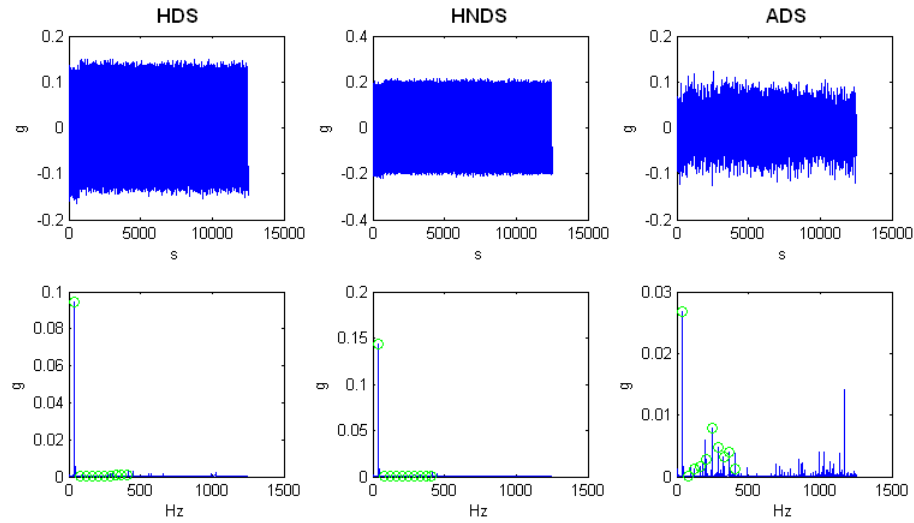
The input of the network consisted of a sum of features computed at each location. Thus while computing of 8 harmonics, the input vector consisted of 24 numbers. Before learning the training set was normalized according to formula

$$\vec{X} = \frac{X}{\sqrt{x_1^2 + \dots + x_n^2}}, \quad (23)$$

where X is the input vector.



(a) Misalignment



(b) Unbalance

Fig. 6. Time waveform and spectra of three different conditions

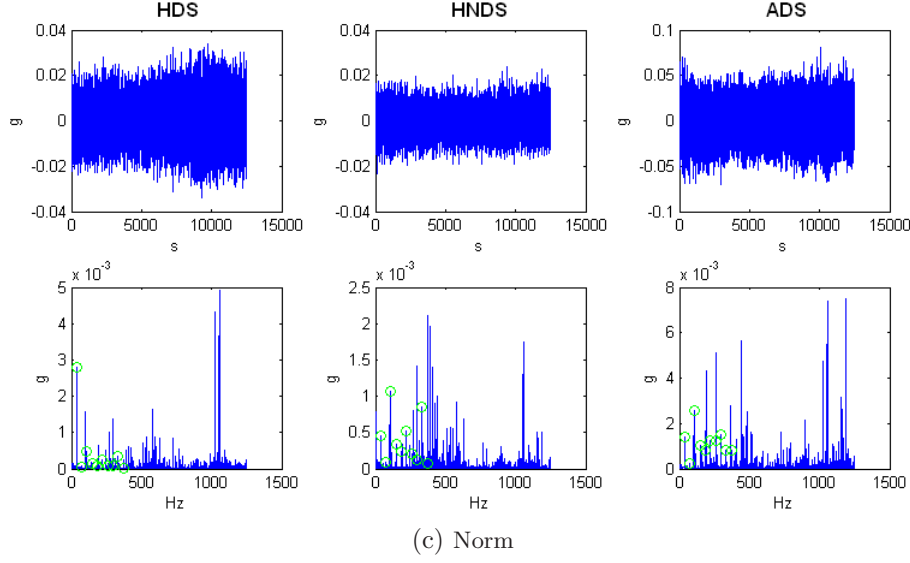


Fig. 6. (Continuation)

4.2. Optimal feature set

The first step was to find the best feature set and compare it with classic methods presented in work [5], which was based on computing of statistical parameters similar like in works [4, 11]. The group of this parameters consists of: root mean square (24), variance (25), kurtosis (26), skewness (27), and normalized six central moment (28).

$$rms = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}, \quad (24)$$

$$\sigma = \mu_2, \quad (25)$$

$$\gamma_3 = \frac{\mu_3}{\sigma^3}, \quad (26)$$

$$\gamma_4 = \frac{\mu_4}{\sigma^4} - 3, \quad (27)$$

$$\gamma_6 = \frac{\mu_6}{\sigma^6}, \quad (28)$$

where μ_i is a central moment of i order (1). Such feature set was found to be enough to achieve 100% diagnosis of the machine with faulty bearing. Also

Tab. 6. The comparison of the feature extraction

Locations	Features	Correct cls./total cls.	Performance
1,2,3	$rms, \sigma^2, \gamma_3, \gamma_4, \gamma_6$	121/151	80,1%
1,2,3	$rms, \sigma^2, \gamma_3, \gamma_4$	121/151	80,1%
1,2,3	rms, σ^2, γ_3	112/151	74,2%
1,2,3	S_{10}	151/151	100,0%
1,2,3	S_9	151/151	100,0%
1,2,3	S_8	151/151	100,0%
1,2,3	S_7	151/151	100,0%
1,2,3	S_6	150/151	99,3%
1,2,3	S_5	145/151	96,0%
1,2,3	S_4	143/151	94,7%
1,2,3	S_3	149/151	98,7%
1,2,3	S_2	141/151	93,4%
1,2,3	$rms, \sigma^2, \gamma_3, \gamma_4, S_5$	121/151	80,1%
1,2,3	$rms, \sigma^2, \gamma_3, S_5$	113/151	74,8%
1,2,3	rms, σ^2, S_5	141/151	93,4%

Tab. 7. The use of different locations

Locations	Features	Correct cls./total cls.	Performance
1,2,3	S_8	151/151	100,0%
1,2	S_8	146/151	96,6%
2,3	S_8	147/151	97,3%
1,3	S_8	150/151	99,3%

the investigations with use of genetic algorithms (Section 2.2.) have proved that increasing statistic parameters does not improve the performance of a network. Tab. 6 depicts different sets of features tested on a probabilistic neural network. All test were taken with data from three locations. It can be seen from the Tab. 7 that harmonic parameters are much better than statistical. The best configuration consists form 7 to 10 harmonics and has achieved 100% of good classification in comparison with 80,1% with statistical features.

It was also found that removing one of the vertical measurements form the network input did not decrease significantly the performance of the network. However, the biggest change was when axial measurement had been removed. These results show that axial measurement contains different information than vertical and it is important in a proper fault detection. In fact it was widely recognized that such defect like angular misalignment manifests with higher axial vibrations.

Tab. 8. Different configurations of the MLP

Network structure	Epochs	Error	Time (s)	Correct cls./total cls.	Performance
10,25,3	6000	0.021213	43.078	147/151	97,3%
10,25,3	9000	0.020331	65.953	147/151	97,3%
10,25,3	3000	0.033073	22.032	145/151	96,0%
10,20,3	3713	0.025000	23.938	145/151	96,0%
10,15,3	6000	0.025726	34.594	143/151	94,7%
5,15,3	6000	0.025454	30.656	145/151	96,0%
15,3	6693	0.024999	28.093	145/151	96,0%
15,3	8000	0.026436	32.718	145/151	96,0%
10,3	7000	0.028174	24.266	145/151	96,0%
10,3	9000	0.026559	29.734	145/151	96,0%

4.3. MLP

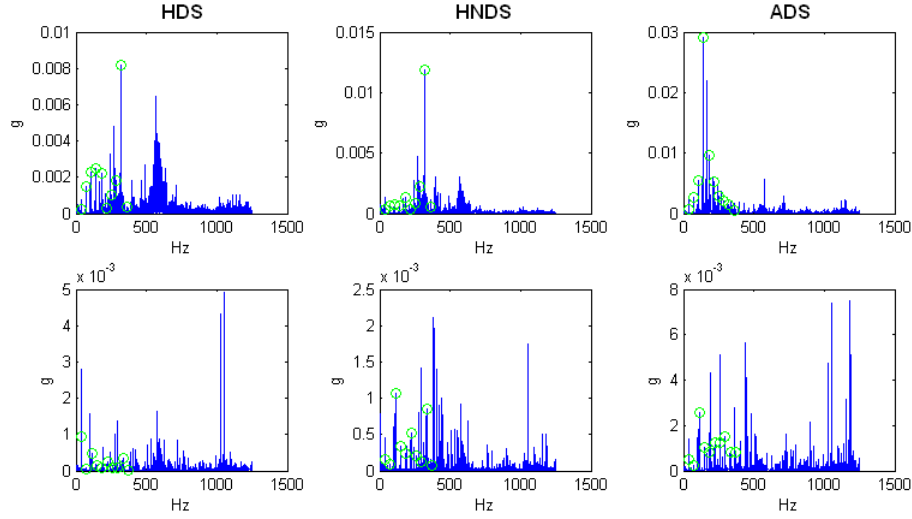
The crucial test was the comparison of a probabilistic neural network with a multi-layered perceptron. The MPL is a network whose abilities in vibrodiagnosis problem were checked and described in many publications [4, 5, 6]. The training and testing set were the same as in PPN testing. The network was trained with backpropagation algorithm with a momentum rule and batch mode. The Tab. 8 contains results of the MPL test with different configurations. The first column contains the number of neurons in particular layers beginning from the hidden layer to the output layer. As we can see, there have been two-layered as well as three-layered structures tested. In all cases the output of the network was a tree-dimensional vector whose one in the n -th place were representing detection of the n -th class. Each neuron had a logistic function (29).

$$\varphi(x) = \frac{1}{1 + \exp(-ax)}. \quad (29)$$

The input vector of the network consisted of eight harmonics from each location. So it was the set of features for which the probabilistic neural network has achieved the best performance. It can be seen from Tab. 9 that the network with two hidden layers were enough to achieve a satisfactory performance of 97,3%. However, it is worse by 2,7% from PNN network. All examples which have been badly recognized by MLP contained good condition and have been classified as misalignment. In Fig. 6 we have depicted comparison of such one bad-classified example (top) with other good condition which was classified correctly (down) . We can clearly see that the first measurement has higher vibrations in all directions especially 9x harmonic

Tab. 9. The MLP with statistical features ($rms, \sigma^2, \gamma_3, \gamma_4, \gamma_6$)

Network structure	Epochs	Error	Time (s)	Correct cls./total cls.	Performance
10,25,3	3000	0.070600	20.859	131/151	86,7%
10,25,3	500	0.094691	3.391	131/151	86,7%
10,20,3	5000	0.063211	34.360	134/151	88,7%
10,15,3	7000	0.056743	47.703	135/151	89,4%

**Fig. 7.** A comparison of bad and good-classified normal conditions

in HDS and HNDS locations are very big. The cause of this strange spectra isn't known and could be an effect of invalid measurement. However, it should be noticed that PPN network had coped with these cases.

Implementations of all networks and algorithms were done in Matlab 7 (R14) environment, on the computer with Intel Centrino Core Duo T2400 1.83 GHz processor and 1.00 GB DDR2 memory.

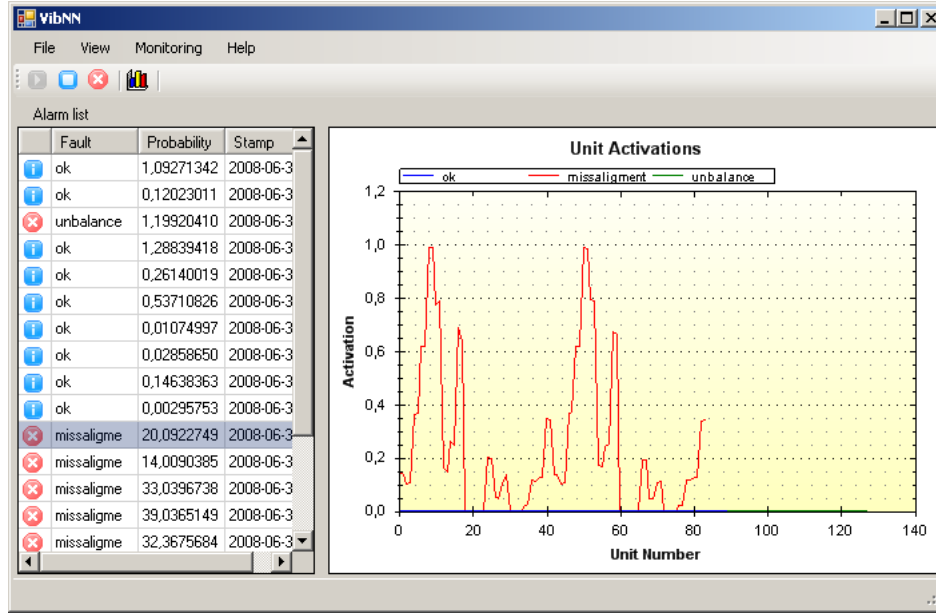


Fig. 8. The main view of the VibNN

5. VibNN

The VibNN is an application which classifies the machine condition by using probabilistic neural networks. The structure of the network was modified so that it could recognize defects that did not occur in the training set as it was described in section In Fig. 8 we can see the main dialog of the application. As we could see except from diagnosis and sum of the probabilities related with them an operator can investigate a respective units activation. This information can be useful while examination fault detection because the same sum of probability can consist of one big peak or many small peaks. On such plot the similarity to other faults can also be analyzed.

6. Concluding remarks

It was found that the probabilistic neural network is a good tool for the machine fault detection, especially when we combine it with the extraction of harmonic features. PPN is better than a multi layered perceptron because it has achieved better performance and it does not consume time for learning.

Furthermore the proposed system has ability of detecting defects that did not occur in the training set and a maintenance engineer can use *a priori* probabilities of fault occurrence for better diagnosis. The minus of such a network can be big memory requirements, but nowadays it seems to be a rather small problem.

In the future work the use of system consisting of two probabilistic networks can be investigated. One network will be responsible for detection of such faults as misalignment and unbalance and will be using harmonic features as in this work. The other will be responsible for detection faulty bearings. The feature set can consist of amplitudes of characteristic bearing faults frequencies and additionally statistical features as it is known that bearing faults manifest in time series. Generalization ability and working with real data of probabilistic neural networks should be also checked. Another interesting thing is investigating how using phase shift measurements between different locations and phase from FFT algorithm will influence the detection process.

7. References

- [1] Reliability direct, <http://www.reliabilitydirect.com>. [22]
- [2] van der Merwe N.T., Hoffman A.J.; *The application of neural networks to vibrational diagnostics for multiple fault conditions*, Computer Standards & Interfaces, 24, 2002. [27]
- [3] Ebersbach S., and Peng Z.; *Expert system development for vibration analysis in machine condition Monitoring*, Expert Systems with Applications, 34, 2008. [25]
- [4] McCormick A.C., Nandi A.K.; *Classification of the rotating machine condition using artificial neural networks*, Proceedings of the IMechE, 1997. [26, 32, 36, 38, 40]
- [5] Samanta B., Al-Balushi K.R.; *Artificial neural network based fault diagnostics of rolling element bearings using time-domain features*, Mechanical Systems and Signal Processing, 17(2), 2001. [24, 32, 36, 38, 40]
- [6] Samanta B., Al-Balushi K.R., Al-Araimi S.A.; *Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection*, Engineering Applications of Artificial Intelligence, 16, 2003. [26, 32, 40]
- [7] Specht D.F.; *Probabilistic neural networks for classification, mapping, or associative memory*, Proceedings of the IEEE International Conference on Neural Networks, 1, 1988. [28]

- [8] Specht D.F.; *Probabilistic neural networks*, Neural Networks, 3, 1990. [28]
- [9] Wang C.-C., Too G.-P.-J., *Rotating machine fault detection based on hos and artificial neural Networks*, Journal of Intelligent Manufacturing, 2002. [28]
- [10] Yanga B.-S., Limb D.-S., Tan A.C.C.; *Vibex: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table monitoring*, Expert Systems with Applications, 28, 2005. [25]
- [11] Zhang L., Lindsay B.J., Nandi A.K.; *Fault detection using genetic programming*, Mechanical Systems and Signal Processing, 19, 2005. [26, 38]
- [12] Zhong B., MacIntyre J., He1 Y., and Tait J.; *High order neural networks for simultaneous diagnosis of multiple faults in rotating machines*, Neural Computing & Applications, 8, 1999. [27]

Received March 15, 2008